

# Information bounds for Gaussian copulas

Peter D. Hoff<sup>1</sup>, Xiaoyue Niu<sup>2</sup>, and Jon A. Wellner<sup>1</sup>

<sup>1</sup>Departments of Statistics and Biostatistics, University of Washington

<sup>2</sup>Department of Statistics, Penn State University

October 18, 2011

## Abstract

Often of primary interest in the analysis of multivariate data are the copula parameters describing the dependence among the variables, rather than the univariate marginal distributions. Since the ranks of a multivariate dataset are invariant to changes in the univariate marginal distributions, rank-based procedures are natural candidates as semiparametric estimators of copula parameters. Asymptotic information bounds for such estimators can be obtained from an asymptotic analysis of the rank likelihood, i.e. the probability of the multivariate ranks. In this article, we obtain limiting normal distributions of the rank likelihood for Gaussian copula models. Our results cover models with structured correlation matrices, such as exchangeable, autoregressive and circular correlation, as well as unstructured correlation matrices. For all Gaussian copula models, the limiting distribution of the rank likelihood ratio is shown to be equal to that of a parametric likelihood ratio for an appropriately chosen multivariate normal model. This implies that the semiparametric information bounds for rank-based estimators are the same as the information bounds for estimators based on the full data, and that the multivariate normal distributions are least favorable.

*Some key words:* copula model, local asymptotic normality, multivariate rank statistics, marginal likelihood, rank likelihood, transformation model.

---

Address for correspondence: [pdhoff@uw.edu](mailto:pdhoff@uw.edu). Peter Hoff's research was supported in part by NI-CHD grant 1R01HD067509-01A1. Jon Wellner's research was supported in part by NSF Grants DMS-0804587 and DMS-1104832, by NI-AID grant 2R01 AI291968-04, and by the Alexander von Humboldt Foundation.

# 1 Rank likelihood for copula models

Recall that a copula is a multivariate CDF having uniform univariate marginal distributions. For any multivariate CDF  $F(y_1, \dots, y_p)$  with continuous margins  $F_1, \dots, F_p$ , the corresponding copula  $C(u_1, \dots, u_p)$  is given by

$$C(u_1, \dots, u_p) = F(F_1^{-1}(u_1), \dots, F_p^{-1}(u_p)).$$

Sklar's theorem [Sklar, 1959] shows that  $C$  is the unique copula for which  $F(y_1, \dots, y_p) = C(F_1(y_1), \dots, F_p(y_p))$ .

In this article we consider models consisting of multivariate probability distributions for which the copula is parameterized separately from the univariate marginal distributions. Specifically, the models we consider consist of collections of multivariate CDFs  $\{F(\mathbf{y}|\theta, \psi) : \mathbf{y} \in \mathbb{R}^p, (\theta, \psi) \in \Theta \times \Psi\}$  such that  $\psi$  parameterizes the univariate marginal distributions and  $\theta$  parameterizes the copula, meaning that for a random vector  $\mathbf{Y} = (Y_1, \dots, Y_p)^T$  with CDF  $F(\mathbf{y}|\theta, \psi)$ ,

$$\begin{aligned} \Pr(Y_j \leq y_j | \theta, \psi) &= F_j(y_j | \psi) \quad \forall \theta \in \Theta, j = 1, \dots, p \\ \Pr(F_1^{-1}(Y_1 | \psi) \leq u_1, \dots, F_p^{-1}(Y_p) \leq u_p | \theta, \psi) &= C(u_1, \dots, u_p | \theta) \quad \forall \psi \in \Psi. \end{aligned}$$

We refer to such a class of distributions as a *copula-parameterized model*. For such a model, it will be convenient to refer to the class of copulas  $\{C(\mathbf{u}|\theta) : \theta \in \Theta\}$  as the copula model, and the class  $\{F_1(y|\psi), \dots, F_p(y|\psi) : \psi \in \Psi\}$  as the marginal model.

As an example, the copula model for the class of  $p$ -variate multivariate normal distributions is called the Gaussian copula model, and is parameterized by letting  $\Theta$  be the set of  $p \times p$  correlation matrices. The marginal model for the  $p$ -variate normal distributions is the set of all  $p$ -tuples of univariate normal distributions. The copula-parameterized models we focus on in this article are semiparametric Gaussian copula models [Klaassen and Wellner, 1997], for which the copula model is Gaussian and the marginal model consists of the set of all  $p$ -tuples of continuous univariate CDFs.

Let  $\mathbf{Y}$  be an  $n \times p$  random matrix whose rows  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d. samples from a  $p$ -variate population. We define the multivariate rank function  $R(\mathbf{Y}) : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^{n \times p}$  so that  $R_{i,j}$ , the  $i, j$ th element of  $R(\mathbf{Y})$ , is the rank of  $Y_{i,j}$  among  $\{Y_{1,j}, \dots, Y_{n,j}\}$ . Note that the ranks  $R(\mathbf{Y})$  are invariant to strictly increasing transformations of the columns of  $\mathbf{Y}$ , and therefore the probability distribution of  $R(\mathbf{Y})$  does not depend on the univariate marginal distributions of the  $p$  variables. As a result, for any copula parameterized model and data

matrix  $\mathbf{y} \in \mathbb{R}^{n \times p}$  with ranks  $R(\mathbf{y}) = \mathbf{r}$ , the likelihood  $L(\theta, \psi : \mathbf{y})$  can be decomposed as

$$\begin{aligned} L(\theta, \psi : \mathbf{y}) = p(\mathbf{y}|\theta, \psi) &= \Pr(R(\mathbf{Y}) = \mathbf{r}|\theta, \psi) \times p(\mathbf{y}|\theta, \psi, \mathbf{r}) \\ &\equiv L(\theta : \mathbf{r}) \times L(\theta, \psi : [\mathbf{y}|\mathbf{r}]), \end{aligned} \quad (1)$$

where  $p(\mathbf{y}|\theta, \psi)$  is the joint density of  $\mathbf{Y}$  and  $p(\mathbf{y}|\theta, \psi, \mathbf{r})$  is the conditional density of  $\mathbf{Y}$  given  $R(\mathbf{Y}) = \mathbf{r}$ . The function  $L(\theta : \mathbf{r}) = \Pr(R(\mathbf{Y}) = \mathbf{r}|\theta)$  is called the *rank likelihood function*. In situations where  $\theta$  is the parameter of interest and  $\psi$  a nuisance parameter, the rank likelihood function can be used to obtain estimates of  $\theta$  without having to estimate the margins or specify a marginal model. A univariate rank likelihood function was proposed by Pettitt [1982] for estimation in monotonically transformed regression models. Asymptotic properties of the rank likelihood for this regression model were studied by Bickel and Ritov [1997], and a parameter estimation scheme based on Gibbs sampling was provided in Hoff [2008]. Rank likelihood estimation of copula parameters was studied in Hoff [2007], who also extended the rank likelihood to accommodate multivariate data with mixed continuous and discrete marginal distributions.

The rank likelihood is constructed from the marginal probability of the ranks and can therefore be viewed as a type of marginal likelihood. Marginal likelihood procedures are often used for estimation in the presence of nuisance parameters (see Section 8.3 of Severini [2000] for a review). Ideally, the statistic that generates a marginal likelihood is “partially sufficient” in the sense that it contains all of the information about the parameter of interest that can be quantified without specifying the nuisance parameter. Notions of partial sufficiency include  $G$ -sufficiency [Barnard, 1963] and  $L$ -sufficiency [Rémon, 1984], which are motivated by group invariance and profile likelihood, respectively. Hoff [2007] showed that the ranks  $R(\mathbf{Y})$  are both a  $G$ - and  $L$ -sufficient statistic in the context of copula estimation.

Although rank-based estimators of the copula parameter  $\theta$  may be appealing for the reasons described above, one may wonder to what extent they are efficient. The decomposition given in (1) indicates that rank-based estimates do not use any information about  $\theta$  contained in  $L(\theta, \psi : [\mathbf{y}|\mathbf{r}])$ , i.e. the conditional probability of the data given the ranks. For at least one copula model, this information is asymptotically negligible: Klaassen and Wellner [1997] showed that for the bivariate normal copula model, a rank-based estimator is semiparametrically efficient and has asymptotic variance equal to the Cramér-Rao information bound in the bivariate normal model, i.e. the bivariate normal model is the least favorable submodel. Genest and Werker [2002] studied the efficiency properties of pseudo-likelihood estimators for two-dimensional semiparametric copula models and show that the pseudo-likelihood estimators (which are functions of the bivariate ranks) are not in general

semiparametrically efficient for non-Gaussian copulas. Chen et al. [2006] proposed estimators in general multivariate copula models that achieve semiparametric asymptotic efficiency but are not based solely on the multivariate ranks. It remains unclear whether estimators based solely on the ranks can be asymptotically efficient in general semiparametric copula models. In particular, it is not yet known if maximum likelihood estimators based on rank likelihoods for Gaussian semiparametric copula models are semiparametrically efficient.

The potential efficiency loss of rank-based estimators can be investigated via the limiting distribution of an appropriately scaled rank likelihood ratio. Generally speaking, the local asymptotic normality (LAN) of a likelihood ratio plays an important role in the asymptotic analysis of testing and estimation procedures. For semiparametric models, the asymptotic variance of a LAN likelihood ratio can be related to efficient tests [Choi et al., 1996] and information bounds for regular estimators [Begun et al., 1983, Bickel et al., 1993]. In particular, the variance of the limiting normal distribution of a LAN rank likelihood ratio provides information bounds for locally regular rank-based estimators of copula parameters.

In this article we obtain the limiting normal distributions of the rank likelihood ratio for Gaussian copula models with structured and unstructured correlation matrices. In the next section we develop several lemmas that give sufficient conditions under which the rank likelihood is LAN. The basic result is that the rank likelihood is LAN if there exists a good rank-measurable approximation to a LAN submodel. For Gaussian copulas, the natural candidate submodels are multivariate normal models, for which the likelihood is quadratic in the observations. In Section 3, we prove a theorem that identifies the types of normal quadratic forms that have good rank-measurable approximations. This result allows us to identify multivariate normal submodels with likelihood ratios that asymptotically approximate the rank likelihood ratio. In Section 4, this result is applied to a class of Gaussian copula models for which the rows of the correlation matrices are permutations of each other. This class includes exchangeable, block-exchangeable and circular correlation models, among others. For a model in this class, the limiting distribution of the rank likelihood ratio is shown to be the same as that of the parametric likelihood ratio for the corresponding multivariate normal model having equal marginal variances. More generally, in Section 5 we show that for any smoothly parameterized Gaussian copula, the rank likelihood ratio is LAN with an asymptotic variance equal to that of the likelihood ratio for the corresponding multivariate normal model with unequal marginal variances. Since the parametric multivariate normal model is a submodel of the semiparametric Gaussian copula model, and in general the semiparametric information bound based on the full data is higher than that of any parametric submodel,

our results imply that the bounds for rank-based estimators are equal to the semiparametric bounds for estimators based on the full data, and that the multivariate normal models are least favorable. This is discussed further in Section 6.

## 2 Approximating the rank likelihood ratio

The local log rank likelihood ratio is defined as

$$\lambda_r(s) = \log \frac{L(\theta + s/\sqrt{n} : \mathbf{r})}{L(\theta : \mathbf{r})},$$

where  $L(\theta : \mathbf{r})$  is defined in (1). Studying  $\lambda_r$  is difficult because  $L(\theta : \mathbf{r})$  is the integral of a copula density over a complicated set defined by multivariate order constraints. However, in some cases it is possible to obtain the asymptotic distribution of  $\lambda_r$  by relating it to the local log likelihood ratio  $\lambda_y$  of an appropriate parametric multivariate model. In this section, we will show that if we can find a sufficiently good rank-measurable approximation  $\lambda_{\hat{y}}$  of  $\lambda_y$ , then the limiting distribution of  $\lambda_r$  will match that of  $\lambda_y$ . This is analogous to the approach taken by Bickel and Ritov [1997] in their investigation of the rank likelihood ratio for a univariate semiparametric regression model.

Define the local log likelihood ratio of a copula parameterized model as

$$\lambda_y(s, t) = \log \frac{L(\theta + s/\sqrt{n}, \psi + t/\sqrt{n} : \mathbf{y})}{L(\theta, \psi : \mathbf{y})}, \quad (2)$$

where  $L(\theta, \psi : \mathbf{y})$  is the (parametric) likelihood function for a given dataset  $\mathbf{y} \in \mathbb{R}^{n \times p}$ . The lack of dependence of the rank likelihood on the marginal distributions leads to the following identity relating  $\lambda_r(s)$  to  $\lambda_y(s, t)$ :

**Lemma 2.1.** *Let  $\mathbf{Y}$  be a random data matrix with rows  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \stackrel{\text{iid}}{\sim} F(\mathbf{y}|\theta, \psi)$ . For every value of  $\psi$ ,  $s$  and  $t$  such that  $\theta, \theta + s/\sqrt{n} \in \Theta$  and  $\psi, \psi + t/\sqrt{n} \in \Psi$ ,*

$$\lambda_r = \log \mathbb{E}[e^{\lambda_y} | R(\mathbf{Y}) = \mathbf{r}].$$

*Proof.*

$$\begin{aligned} \log \mathbb{E}[e^{\lambda_y} | R(\mathbf{Y}) = \mathbf{r}] &= \log \int_{R(\mathbf{y})=\mathbf{r}} \frac{p(\mathbf{y}|\theta + s/\sqrt{n}, \psi + t/\sqrt{n})}{p(\mathbf{y}|\theta, \psi)} \frac{p(\mathbf{y}|\theta, \psi)}{\Pr(R(\mathbf{Y}) = \mathbf{r}|\theta)} d\mathbf{y} \\ &= \log \frac{\Pr(R(\mathbf{Y}) = \mathbf{r}|\theta + s/\sqrt{n})}{\Pr(R(\mathbf{Y}) = \mathbf{r}|\theta)} = \lambda_r. \end{aligned}$$

□

Note that the identity holds for any marginal model and for any value of  $\psi$  and  $t$  for which  $\psi$  and  $\psi + t/\sqrt{n}$  are in the parameter space. Therefore, the numerical value of  $\lambda_r$  is invariant to the marginal distribution under which it is calculated.

Now suppose we would like to describe the statistical properties of  $\lambda_r$  when the matrix  $\mathbf{r}$  is replaced by the ranks  $R(\mathbf{Y})$ , where the rows of  $\mathbf{Y}$  are i.i.d. samples from a population with copula  $C(\mathbf{u}|\theta)$ . Since the distribution of the ranks of  $\mathbf{Y}$  is invariant with respect to the marginal distributions, the choice of the marginal model in Lemma 2.1 is immaterial and can be selected to facilitate analysis. Our strategy will be to select a marginal model for which a rank-based approximation  $\lambda_{\hat{y}}$  of  $\lambda_y$  is available. If  $\lambda_{\hat{y}}$  is rank-measurable, then

$$\begin{aligned}\lambda_r &= \log E[e^{\lambda_y} | R(\mathbf{Y})] \\ &= \lambda_{\hat{y}} + \log E[e^{\lambda_y - \lambda_{\hat{y}}} | R(\mathbf{Y})].\end{aligned}$$

If the approximation of  $\lambda_y$  by  $\lambda_{\hat{y}}$  is sufficiently accurate to make the remainder term,  $\log E[e^{\lambda_y - \lambda_{\hat{y}}} | R(\mathbf{Y})]$ , converge in probability to zero as  $n \rightarrow \infty$ , then the asymptotic distribution of  $\lambda_r$  is determined by that of  $\lambda_{\hat{y}}$ . As shown by the following proposition, the local asymptotic normality (LAN) of  $\lambda_y$ , along with  $\lambda_y - \lambda_{\hat{y}} = o_p(1)$ , implies convergence of the remainder. In what follows, all limits are as  $n \rightarrow \infty$  unless otherwise noted.

**Proposition 2.2.** *Let  $\lambda_y$  be LAN and let  $\lambda_{\hat{y}}$  be a rank-measurable approximation to  $\lambda_y$ , such that  $\lambda_y - \lambda_{\hat{y}} \xrightarrow{p} 0$ . If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim i.i.d. F(\mathbf{y}|\theta, \psi)$ , then  $\log E[e^{\lambda_y - \lambda_{\hat{y}}} | R(\mathbf{Y})] \xrightarrow{p} 0$ .*

This proposition was essentially proven at the end of the proof of Theorem 1 of Bickel and Ritov [1997] in the context of the regression transformation model, although details were omitted. We include the proof here for completeness. The proof of Proposition 2.2 makes use of the following lemma about conditional expectations:

**Lemma 2.3.** *If  $E[|X_n|] \rightarrow 0$  and  $Z_n$  is a random sequence, then  $E[X_n | Z_n] \xrightarrow{p} 0$ .*

*Proof.* By Markov's inequality,

$$\begin{aligned}\Pr(|E[X_n | Z_n]| > \epsilon) &\leq E[|E[X_n | Z_n]|] / \epsilon \\ &\leq E[E[|X_n| | Z_n]] / \epsilon \\ &= E[|X_n|] / \epsilon \rightarrow 0.\end{aligned}$$

□

In particular, note that if  $X_n \xrightarrow{p} 0$  and  $X_n$  is bounded or uniformly integrable, then  $E[|X_n|] \rightarrow 0$  and so  $E[X_n | Z_n] \xrightarrow{p} 0$ .

*Proof of proposition 2.2 .* Let  $U_n = e^{\lambda_y}$ ,  $V_n = e^{\lambda_{\hat{y}}}$  and  $\mathbf{R}_n = R(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ , so that the exponential of the remainder term can be written as  $E[\frac{U_n}{V_n} | \mathbf{R}_n]$ . For any  $M > 1$  we can write

$$\begin{aligned}
|E[\frac{U_n}{V_n} - 1 | \mathbf{R}_n]| &\leq E[|\frac{U_n}{V_n} - 1| | \mathbf{R}_n] \\
&= E[|\frac{U_n}{V_n} - 1| 1_{(\frac{U_n}{V_n} \leq M)} | \mathbf{R}_n] + E[|\frac{U_n}{V_n} - 1| 1_{(\frac{U_n}{V_n} > M)} | \mathbf{R}_n] \\
&\leq E[|\frac{U_n}{V_n} - 1| 1_{(\frac{U_n}{V_n} \leq M)} | \mathbf{R}_n] + E[\frac{U_n}{V_n} 1_{(\frac{U_n}{V_n} > M)} | \mathbf{R}_n] \\
&\quad + E[1_{(\frac{U_n}{V_n} > M)} | \mathbf{R}_n] \\
&= E[|\frac{U_n}{V_n} - 1| 1_{(\frac{U_n}{V_n} \leq M)} | \mathbf{R}_n] + V_n^{-1} E[U_n 1_{(\frac{U_n}{V_n} > M)} | \mathbf{R}_n] \\
&\quad + \Pr(\frac{U_n}{V_n} > M | \mathbf{R}_n) \\
&= a_n + b_n + c_n.
\end{aligned}$$

We will show that each of  $a_n$ ,  $b_n$  and  $c_n$  converge in probability to zero. To do so, we will make use of the following facts:

1.  $U_n/V_n = e^{\lambda_y(s) - \lambda_{\hat{y}}(s)} \xrightarrow{p} 1$  by the continuous mapping theorem;
2.  $U_n = e^{\lambda_y(s)}$  and  $V_n^{-1} = e^{-\lambda_{\hat{y}}(s)}$  are bounded in probability, as  $\lambda_y(s)$  and  $\lambda_{\hat{y}}(s)$  converge in distribution.
3.  $\{U_n : n \in \mathbb{N}\}$  is uniformly integrable, since  $\log U_n = \lambda_y(s)$  is LAN [Hall and Loynes, 1977]

To see that  $a_n$  and  $c_n$  converge in probability to zero, note that both  $|\frac{U_n}{V_n} - 1| 1_{(\frac{U_n}{V_n} \leq M)}$  and  $1_{(\frac{U_n}{V_n} > M)}$  are bounded random variables that converge in probability to zero, so their conditional expectations given  $\mathbf{R}_n$  converge in probability to zero by the lemma.

For the sequence  $b_n$ , note that  $U_n$  is  $O_p(1)$  as it converges in distribution, and  $1_{(\frac{U_n}{V_n} > M)}$  is  $o_p(1)$  as  $\frac{U_n}{V_n} \xrightarrow{p} 1$ , so  $\tilde{U}_n = U_n 1_{(\frac{U_n}{V_n} > M)}$  is  $o_p(1)$ . Now  $0 \leq \tilde{U}_n \leq U_n$  for each  $n$ , and  $\{U_n : n \in \mathbb{N}\}$  is uniformly integrable, so  $\{\tilde{U}_n : n \in \mathbb{N}\}$  is uniformly integrable as well. This and  $\tilde{U}_n \xrightarrow{p} 0$  imply that  $E[|\tilde{U}_n|] = E[\tilde{U}_n] \rightarrow 0$ , and so  $E[\tilde{U}_n | \mathbf{R}_n] \xrightarrow{p} 0$  by the lemma. Since  $b_n = V_n^{-1} E[\tilde{U}_n | \mathbf{R}_n]$ , and  $V_n^{-1}$  is  $O_p(1)$ ,  $b_n$  is  $o_p(1)$ .  $\square$

Now if  $\lambda_y$  is LAN then  $\lambda_y \xrightarrow{d} Z$  where  $Z$  has a normal distribution. If the conditions of Proposition 2.2 hold, then we must also have  $\lambda_{\hat{y}} \xrightarrow{d} Z$  and  $\lambda_r \xrightarrow{d} Z$ . We summarize the results of Lemma 2.1 and Proposition 2.2 with the following theorem:

**Theorem 2.4.** *Let  $\{F(\mathbf{y}|\theta, \psi) : \theta \in \Theta, \psi \in \Psi\}$  be a copula parameterized model where for given values of  $\theta$  and  $s$  there exists values of  $\psi$  and  $t$  such that under i.i.d. sampling from  $F(\mathbf{y}|\theta, \psi)$ ,*

1.  $\lambda_y$  is LAN, so that  $\lambda_y \xrightarrow{d} Z$  and

2. there exists a rank-measurable approximation  $\lambda_{\hat{y}}$  such that  $\lambda_y - \lambda_{\hat{y}} \xrightarrow{p} 0$ .

*Then  $\lambda_r \xrightarrow{d} Z$  as  $n \rightarrow \infty$  under i.i.d. sampling from any population with copula  $C(\mathbf{u}|\theta)$  equal to that of  $F(\mathbf{y}|\theta, \psi)$  and arbitrary continuous marginal distributions.*

Note that under the conditions of the theorem and sampling from  $F(\mathbf{y}|\theta, \psi)$ , the differences between each pair of  $\lambda_y$ ,  $\lambda_{\hat{y}}$  and  $\lambda_r$  converge to zero and all three of these log-likelihood ratios converge to the same normal random variable. If the data are being sampled from a population with the same copula as  $F(\mathbf{y}|\theta, \psi)$  but different margins, then there exists a transformation of the data such that  $F(\mathbf{y}|\theta, \psi)$  is the distribution of the transformed population. Thus, all we need is that conditions 1 and 2 hold for some marginal model and values of  $\psi$  and  $t$ . This is enough to give the convergence of  $\lambda_r$  to a normal random variable under sampling from any population with continuous marginal distributions and the same copula as  $F(\mathbf{y}|\theta, \psi)$ .

### 3 Rank approximations to normal quadratic forms

Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  be i.i.d. samples from a member of a class of mean-zero  $p$ -variate normal distributions indexed by a correlation parameter  $\theta \in \Theta$  and a variance parameter  $\psi \in \Psi$ . As discussed in the next section, the local likelihood ratio  $\lambda_y$  can be expressed as a quadratic function of  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , taking the form

$$\lambda_y(s, t) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{Y}_i^T \mathbf{A} \mathbf{Y}_i \right) + c(\theta, \psi, s, t) + o_p(1)$$

for some matrix  $\mathbf{A}$  which could be a function of  $s$ ,  $t$ ,  $\theta$  and  $\psi$ . A natural rank-based approximation to  $\lambda_y$  is

$$\lambda_{\hat{y}}(s, t) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\mathbf{Y}}_i^T \mathbf{A} \hat{\mathbf{Y}}_i \right) + c(\theta, \psi, s, t),$$

where  $\{\hat{Y}_{i,j} : i \in \{1, \dots, n\}, j \in \{1, \dots, p\}\}$  are the (approximate) normal scores, defined by  $\mathbf{R} = R(\mathbf{Y})$  and  $\hat{Y}_{i,j} = \sqrt{\text{Var}[Y_{i,j}|\psi]} \times \Phi^{-1}\left(\frac{R_{i,j}}{n+1}\right)$ . Whether or not  $\lambda_{\hat{y}} - \lambda_y \rightarrow 0$  therefore



depends on the convergence to zero of the difference between the quadratic terms of  $\lambda_{\hat{y}}$  and  $\lambda_y$ .

Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim \text{i.i.d. } N_p(\mathbf{0}, \mathbf{C})$ , where  $\mathbf{C}$  is a correlation matrix, so the normal scores are given by  $\hat{Y}_{i,j} = \Phi^{-1}(\frac{R_{i,j}}{n+1})$ . In this section, we will find conditions on  $\mathbf{C}$  and  $\mathbf{A}$  such that

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \hat{\mathbf{Y}}_i^T \mathbf{A} \hat{\mathbf{Y}}_i - \mathbf{Y}_i^T \mathbf{A} \mathbf{Y}_i \right) \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

Let  $\tilde{\mathbf{A}} = (\mathbf{A} + \mathbf{A}^T)/2$ , so that  $\mathbf{y}^T \tilde{\mathbf{A}} \mathbf{y} = \mathbf{y}^T \mathbf{A} \mathbf{y}$  for all  $\mathbf{y} \in \mathbb{R}^p$ . Then

$$\begin{aligned} \hat{\mathbf{Y}}^T \mathbf{A} \hat{\mathbf{Y}} - \mathbf{Y}^T \mathbf{A} \mathbf{Y} &= \hat{\mathbf{Y}}^T \tilde{\mathbf{A}} \hat{\mathbf{Y}} - \mathbf{Y}^T \tilde{\mathbf{A}} \mathbf{Y} \\ &= (\hat{\mathbf{Y}} - \mathbf{Y})^T \tilde{\mathbf{A}} (\hat{\mathbf{Y}} - \mathbf{Y}) + 2(\hat{\mathbf{Y}} - \mathbf{Y})^T \tilde{\mathbf{A}} \mathbf{Y}, \end{aligned}$$

the latter equality holding since  $\tilde{\mathbf{A}}$  is symmetric. From this, we can write  $S_n = Q_n + 2L_n$  where

$$\begin{aligned} Q_n &= \frac{1}{\sqrt{n}} \sum (\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^T \tilde{\mathbf{A}} (\hat{\mathbf{Y}}_i - \mathbf{Y}_i) \\ L_n &= \frac{1}{\sqrt{n}} \sum (\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^T \tilde{\mathbf{A}} \mathbf{Y}_i. \end{aligned}$$

We will show that  $Q_n \xrightarrow{p} 0$  for all  $\mathbf{C}$  and  $\tilde{\mathbf{A}}$ , and find conditions on  $\mathbf{C}$  and  $\tilde{\mathbf{A}}$  under which  $L_n \xrightarrow{p} 0$ . To do this we make use of a theorem of de Wet and Venter [1972]:

**Theorem** (de Wet and Venter). *Let  $Z_1, \dots, Z_n$  be i.i.d. standard normal variables. Let  $R_i$  be the rank of  $Z_i$ , let  $\hat{Z}_i = \Phi^{-1}[R_i/(n+1)]$ , and let  $W_n = \sum (Z_i - \hat{Z}_i)^2$ . Then*

$$W_n - v_n \xrightarrow{d} \gamma,$$

where  $v_n$  is a sequence of constants such that  $C_1 \log \log n < v_n < C_2 \log \log n$  for all  $n$  and some  $C_1, C_2 > 0$ , and  $\mathbb{E}[\gamma] = 0$  and  $\text{Var}[\gamma] = \pi^2/3$ .

From this theorem we have the following corollary:

**Corollary 3.1.** *Let  $Z_1, \dots, Z_n$  be i.i.d. standard normal random variables. Then  $\sum (Z_i - \hat{Z}_i)^2 / \log \log n$  is bounded in probability, and  $\sum (Z_i - \hat{Z}_i)^2 / \sqrt{n} \xrightarrow{p} 0$ .*

*Proof.* Let  $\tilde{W}_n = W_n / (\log \log n)$  and  $\tilde{v}_n = v_n / (\log \log n)$ , where  $W_n$  and  $v_n$  are defined as in the theorem. Then  $|\tilde{W}_n| \leq |\tilde{W}_n - \tilde{v}_n| + |\tilde{v}_n|$ . Now  $W_n - v_n$  converges in distribution so it is  $O_p(1)$ , and so  $\tilde{W}_n - \tilde{v}_n$  is  $o_p(1)$ . The deterministic term  $\tilde{v}_n$  is bounded by  $C_2$ , and so  $|\tilde{W}_n - \tilde{v}_n| + |\tilde{v}_n|$  is bounded in probability, implying  $|\tilde{W}_n|$  is bounded in probability as well. Convergence in probability of  $\sum (Z_i - \hat{Z}_i)^2 / \sqrt{n}$  to zero then follows.  $\square$

Returning to  $S_n$ , to see that  $Q_n \xrightarrow{p} 0$  note that

$$Q_n = \sum_{j=1}^p \tilde{a}_{j,j} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{Y}_{i,j} - Y_{i,j})^2 \right) + \sum_{j \neq k} \tilde{a}_{j,k} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{Y}_{i,j} - Y_{i,j})(\hat{Y}_{i,k} - Y_{i,k}) \right).$$

The squared terms converge in probability to zero by the corollary, and the cross term converges in probability to zero by the Cauchy-Schwarz inequality.

To find conditions under which  $L_n \xrightarrow{p} 0$ , note that

$$(\hat{\mathbf{y}} - \mathbf{y})^T \tilde{\mathbf{A}} \mathbf{y} = \sum_{j=1}^p (\hat{y}_j - y_j) \tilde{\mathbf{a}}_j^T \mathbf{y}_i$$

where  $\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_p$  are the rows of  $\tilde{\mathbf{A}}$ . This gives

$$L_n = \sum_{j=1}^p L_{n,j} \equiv \sum_{j=1}^p \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{Y}_{i,j} - Y_{i,j}) \tilde{\mathbf{a}}_j^T \mathbf{Y}_i.$$

Let  $\mathbf{c}_j$  be the  $j$ th row of  $\mathbf{C}$ , the correlation matrix of  $\mathbf{Y}$ . We will show that  $L_{n,j} \xrightarrow{p} 0$  if  $\tilde{\mathbf{a}}_j^T \mathbf{c}_j = 0$  using an argument based on conditional expectations. Considering  $L_{n,1}$  for example, we have

$$\begin{aligned} E[L_{n,1} | Y_{1,1}, \dots, Y_{n,1}] &= \frac{1}{\sqrt{n}} \sum (\hat{Y}_{i,1} - Y_{i,1}) E[\tilde{\mathbf{a}}_1^T \mathbf{Y}_i | Y_{i,1}] \\ &= \frac{1}{\sqrt{n}} \sum (\hat{Y}_{i,1} - Y_{i,1}) \tilde{\mathbf{a}}_1^T \mathbf{c}_1 Y_{i,1} = 0 \end{aligned}$$

if  $\tilde{\mathbf{a}}_j^T \mathbf{c}_j = 0$ . The conditional expectation of  $L_{n,1}^2$  is given by

$$\begin{aligned} E[L_{n,1}^2 | Y_{1,1}, \dots, Y_{n,1}] &= \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,1} - Y_{i,1})^2 E[(\tilde{\mathbf{a}}_1^T \mathbf{Y}_i)^2 | Y_{i,1}] + \\ &\quad \frac{1}{n} \sum \sum_{i_1 \neq i_2} (\hat{Y}_{i_1,1} - Y_{i_1,1})(\hat{Y}_{i_2,1} - Y_{i_2,1}) E[\tilde{\mathbf{a}}_1^T \mathbf{Y}_{i_1} | Y_{i_1,1}] E[\tilde{\mathbf{a}}_1^T \mathbf{Y}_{i_2} | Y_{i_2,1}]. \end{aligned}$$

The expectations in the second sum are both  $\tilde{\mathbf{a}}_1^T \mathbf{c}_1 = 0$ , leaving

$$\text{Var}[L_{n,1} | Y_{1,1}, \dots, Y_{n,1}] = E[L_{n,1}^2 | Y_{1,1}, \dots, Y_{n,1}] = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,1} - Y_{i,1})^2 E[(\tilde{\mathbf{a}}_1^T \mathbf{Y}_i)^2 | Y_{i,1}].$$

The conditional expectation  $E[(\tilde{\mathbf{a}}_1^T \mathbf{Y}_i)^2 | Y_{i,1}]$  can be obtained by noting that if  $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{C})$ , then the conditional distribution of  $\mathbf{Y}$  given  $Y_1$  can be expressed as

$$\mathbf{Y} | Y_1 \stackrel{d}{=} \mathbf{c}_1 Y_1 + \mathbf{G}\boldsymbol{\epsilon},$$

where  $\mathbf{G}\mathbf{G}^T = \mathbf{C} - \mathbf{c}_1\mathbf{c}_1^T$  and  $\boldsymbol{\epsilon}$  is  $p$ -variate standard normal. The desired second moment is then

$$\begin{aligned} \mathbb{E}[(\tilde{\mathbf{a}}_1^T \mathbf{Y})^2 | Y_1] &= \tilde{\mathbf{a}}_1^T \mathbb{E}[\mathbf{Y}\mathbf{Y}^T | Y_1] \tilde{\mathbf{a}}_1 \\ &= \tilde{\mathbf{a}}_1^T \mathbb{E}[Y_1^2 \mathbf{c}_1 \mathbf{c}_1^T + 2Y_1 \mathbf{c}_1 \boldsymbol{\epsilon}^T \mathbf{G}^T + \mathbf{G}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \mathbf{G}^T | Y_1] \tilde{\mathbf{a}}_1 \\ &= (Y_1^2 - 1)(\tilde{\mathbf{a}}_1^T \mathbf{c}_1)^2 + \tilde{\mathbf{a}}_1^T \mathbf{C} \tilde{\mathbf{a}}_1 \end{aligned}$$

which is equal to  $\tilde{\mathbf{a}}_1^T \mathbf{C} \tilde{\mathbf{a}}_1$  under the condition that  $\tilde{\mathbf{a}}_1^T \mathbf{c}_1 = 0$ . Letting  $\gamma_1 = \tilde{\mathbf{a}}_1^T \mathbf{C} \tilde{\mathbf{a}}_1$ , the conditional variance of  $L_{n,1}$  given the observations for the first variate is then

$$\text{Var}[L_{n,1} | Y_{1,1}, \dots, Y_{n,1}] = \frac{\gamma_1}{n} \sum (\hat{Y}_{i,1} - Y_{i,1})^2.$$

Applying Chebyshev's inequality gives

$$\begin{aligned} \Pr(|L_{n,1}| > \epsilon | Y_{1,1}, \dots, Y_{n,1}) &\leq 1 \wedge \text{Var}[L_{n,1} | Y_{1,1}, \dots, Y_{n,1}] / \epsilon^2 \\ &= 1 \wedge \frac{\gamma_1}{\epsilon^2} \frac{\sum (\hat{Y}_{i,1} - Y_{i,1})^2}{n} \\ &= 1 \wedge c_n = \tilde{c}_n. \end{aligned}$$

Now  $c_n \xrightarrow{p} 0$  by Corollary 3.1, and therefore so does  $\tilde{c}_n$ . But as  $\tilde{c}_n$  is bounded, we have  $\mathbb{E}[\tilde{c}_n] \rightarrow 0$ , giving

$$\begin{aligned} \Pr(|L_{n,1}| > \epsilon) &= \mathbb{E}[\Pr(|L_{n,1}| > \epsilon | Y_{1,1}, \dots, Y_{n,1})] \\ &\leq \mathbb{E}[\tilde{c}_n] \rightarrow 0, \end{aligned}$$

and so  $L_{n,1} \xrightarrow{p} 0$ . The same argument can be applied to  $L_{n,j}$  for each  $j$ , and so  $L_n = \sum_{j=1}^p L_{n,j} \rightarrow 0$  as long as  $\tilde{\mathbf{a}}_j^T \mathbf{c}_j = 0$  for each  $j = 1, \dots, p$ , or equivalently, if the diagonal elements of  $\mathbf{A}\mathbf{C} + \mathbf{A}^T \mathbf{C}$  are zero. This result is summarized in the following theorem:

**Theorem 3.2.** *Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim i.i.d. N_p(\mathbf{0}, \mathbf{C})$  where  $\mathbf{C}$  is a correlation matrix, and let  $\hat{Y}_{i,j} = \Phi^{-1}(\frac{R_{i,j}}{n+1})$ , where  $R_{i,j}$  is the rank of  $Y_{i,j}$  among  $Y_{1,j}, \dots, Y_{n,j}$ . Let  $\mathbf{A}$  be a matrix such that  $\text{diag}(\mathbf{A}\mathbf{C} + \mathbf{A}^T \mathbf{C}) = \mathbf{0}$ . Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\mathbf{Y}}_i^T \mathbf{A} \hat{\mathbf{Y}}_i - \mathbf{Y}_i^T \mathbf{A} \mathbf{Y}_i) \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Note that if  $\mathbf{A}$  is symmetric, the condition on  $\mathbf{A}$  reduces to the diagonal elements of  $\mathbf{A}\mathbf{C}$  being zero.

## 4 Submodels having equal marginal variances

### 4.1 The bivariate normal model

Klaassen and Wellner [1997] showed that for the bivariate normal copula model with correlation  $\theta \in [-1, 1]$ , a rank-based estimate (the normal scores rank correlation coefficient) is asymptotically efficient and has asymptotic variance equal to  $(1 - \theta^2)^2$ . That  $(1 - \theta^2)^2$  is a lower bound on the variance of rank-based estimators of  $\theta$  can be easily confirmed using the results of the previous section.

As these authors note, this asymptotic variance is equal to that of the MLE for  $\theta$  in the case of the bivariate normal model with equal but unknown marginal variances, which thus constitutes a least favorable parametric submodel. This suggests that the local likelihood ratio for this submodel provides an asymptotic approximation to the rank likelihood ratio. To verify this, let  $l(\mathbf{y})$  be the log probability density for the class of mean-zero bivariate normal densities with correlation  $\theta$  and equal marginal precisions (inverse-variances)  $\psi$ . The log likelihood derivatives are

$$\begin{aligned} i_\theta &= \frac{\theta(1 - \theta^2) + \psi y_1 y_2 (1 + \theta^2) - \psi(y_1^2 + y_2^2)\theta}{(1 - \theta^2)^2} \\ i_\psi &= \frac{1}{\psi} - \frac{(y_1^2 + y_2^2)/2 - \theta y_1 y_2}{1 - \theta^2}. \end{aligned}$$

The information matrix is

$$I = E[(\nabla l)(\nabla l)^T] = \begin{pmatrix} I_{\theta\theta} & I_{\theta\psi} \\ I_{\psi\theta} & I_{\psi\psi} \end{pmatrix} = \begin{pmatrix} \frac{1+\theta^2}{(1-\theta^2)^2} & \frac{\theta}{\psi(1-\theta^2)} \\ \frac{\theta}{\psi(1-\theta^2)} & \frac{1}{\psi^2} \end{pmatrix},$$

which gives the efficient score function

$$i_\theta^* = i_\theta - I_{\theta\psi} I_{\psi\psi}^{-1} i_\psi = \frac{1}{(1 - \theta^2)^2} (y_1 y_2 - \theta(y_1^2 + y_2^2)/2),$$

the efficient influence function

$$\tilde{l}_\theta = y_1 y_2 - \theta(y_1^2 + y_2^2)/2,$$

and the efficient information

$$\begin{aligned} I_{\theta\theta \cdot \psi} &= ([I^{-1}]_{11})^{-1} = I_{\theta\theta} - I_{\theta\psi}^2 I_{\psi\psi}^{-1} \\ &= \frac{1 + \theta^2}{(1 - \theta^2)^2} - \frac{\theta^2}{(1 - \theta^2)^2} = \frac{1}{(1 - \theta^2)^2}, \end{aligned}$$

and so  $I_{\theta\theta\cdot\psi}^{-1} = (1 - \theta^2)^2$  is the information bound for estimators of  $\theta$  in this parametric model.

If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d.  $N_2(\mathbf{0}, \mathbf{C}(\theta)/\psi)$ , where  $\mathbf{C}(\theta) = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ , then the local likelihood ratio  $\lambda_y(s, t)$  defined by (2) has a limiting normal distribution:

$$\begin{aligned} \lambda_y(s, t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [s \times \dot{l}_\theta(\mathbf{Y}_i) + t \times \dot{l}_\psi(\mathbf{Y}_i)] - [s, t] I[s, t]^T / 2 + o_p(1) \\ &\xrightarrow{d} N(-[s, t] I[s, t]^T / 2, [s, t] I[s, t]^T) \text{ as } n \rightarrow \infty. \end{aligned}$$

Letting  $t = -I_{\psi\psi}^{-1} I_{\psi\theta} s$ , the local log likelihood ratio  $\lambda_y(s)$  can be expressed as

$$\lambda_y(s) = \frac{s}{\sqrt{n}} \sum [\dot{l}_\theta(\mathbf{Y}_i) - I_{\theta\psi} I_{\psi\psi}^{-1} \dot{l}_\psi(\mathbf{Y}_i)] - \frac{1}{2} s^2 I_{\theta\theta\cdot\psi} + o_p(1),$$

which converges as  $n \rightarrow \infty$  to a  $N(-s^2 I_{\theta\theta\cdot\psi}/2, s^2 I_{\theta\theta\cdot\psi})$  random variable if  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d.  $N_2(\mathbf{0}, \mathbf{C}(\theta)/\psi)$ . Note that  $I_{\theta\theta\cdot\psi}$ , and therefore the asymptotic distribution of  $\lambda_y$ , do not depend on the value of  $\psi$ . Letting  $\psi = 1$ , the value of  $\dot{l}_\theta - I_{\theta\psi} I_{\psi\psi}^{-1} \dot{l}_\psi$  is given by

$$\begin{aligned} \dot{l}_\theta(\mathbf{y}) - I_{\theta\psi} I_{\psi\psi}^{-1} \dot{l}_\psi(\mathbf{y}) &= \frac{y_1 y_2 - \theta(y_1^2 + y_2^2)/2}{(1 - \theta^2)^2} \\ &= \mathbf{y}^T \mathbf{A} \mathbf{y}, \end{aligned}$$

where  $\mathbf{A}$  is the symmetric matrix

$$\mathbf{A} = \frac{1}{2(1 - \theta^2)^2} \begin{bmatrix} -\theta & 1 \\ 1 & -\theta \end{bmatrix}.$$

We are now in a position to apply the results of the previous sections. If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim$  i.i.d.  $N_2(\mathbf{0}, \mathbf{C}(\theta))$ , then  $\lambda_y(s)$  is LAN, so that

$$\lambda_y(s) = \frac{s}{\sqrt{n}} \sum \mathbf{Y}_i^T \mathbf{A} \mathbf{Y}_i - s^2 I_{\theta\theta\cdot\psi}/2 + o_p(1) \xrightarrow{d} N(-s^2 I_{\theta\theta\cdot\psi}/2, s^2 I_{\theta\theta\cdot\psi}).$$

A rank-measurable approximation to  $\lambda_y(s)$  is given by

$$\lambda_y(s) = \frac{s}{\sqrt{n}} \sum \hat{\mathbf{Y}}_i^T \mathbf{A} \hat{\mathbf{Y}}_i - s^2 I_{\theta\theta\cdot\psi}/2.$$

Letting  $\mathbf{C}$  be the bivariate correlation matrix with correlation  $\theta$ , it is easy to check that the diagonal elements of  $\mathbf{A}\mathbf{C}$  are zero. Since  $\mathbf{A}$  is also symmetric, the conditions of Theorem 3.2 are met and we have  $\lambda_{\hat{y}} - \lambda_y = o_p(1)$ . By Theorem 2.4, under i.i.d. sampling from a population with a bivariate normal copula with correlation  $\theta$ , we have

$$\lambda_r(s) \xrightarrow{d} N(-s^2 I_{\theta\theta\cdot\psi}/2, s^2 I_{\theta\theta\cdot\psi}).$$

The semiparametric information bound for rank-based estimators of  $\theta$  is thus  $I_{\theta\theta\cdot\psi}^{-1}$ , confirming the result of Klaassen and Wellner [1997].

## 4.2 Multivariate models

A natural question is whether or not the above result can be extended to  $p$ -variate Gaussian copula models where  $p > 2$ . In this section we identify a class of correlation models for which the limiting distribution of the rank likelihood ratio is the same as that of the likelihood ratio for the corresponding parametric multivariate normal model with equal marginal variances.

Let  $\{N_p(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})/\psi) : \boldsymbol{\theta} \in \Theta, \psi \in \mathbb{R}^+\}$  denote a class of  $p$ -variate mean-zero normal distributions, where  $\psi$  parameterizes the common marginal precision and the correlation matrix is parameterized as  $\mathbf{C}(\boldsymbol{\theta})$ , a twice-differentiable matrix-valued function of  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ . For the calculations that follow, it will be convenient to express the likelihood in terms of the inverse correlation matrix  $\mathbf{B}(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})^{-1}$ , giving

$$l(\mathbf{y} : \boldsymbol{\theta}, \psi) = (-p \log 2\pi + p \log \psi + \log |\mathbf{B}| - \psi \mathbf{y}^T \mathbf{B} \mathbf{y})/2. \quad (3)$$

The corresponding likelihood derivatives are

$$\begin{aligned} \dot{l}_\psi(\mathbf{y}) &= (p/\psi - \mathbf{y}^T \mathbf{B} \mathbf{y})/2 & \ddot{l}_{\theta_k \psi}(\mathbf{y}) &= -\mathbf{y}^T \mathbf{B}_{\theta_k} \mathbf{y}/2 \\ \dot{l}_{\theta_k}(\mathbf{y}) &= (\text{tr}(\mathbf{B}_{\theta_k} \mathbf{C}) - \psi \mathbf{y}^T \mathbf{B}_{\theta_k} \mathbf{y})/2 & \ddot{l}_{\psi \psi}(\mathbf{y}) &= -p/(2\psi^2), \end{aligned}$$

where  $\mathbf{B}_{\theta_k}$  is the matrix of derivatives of the elements of  $\mathbf{B}$  with respect to a particular element  $\theta_k$  of  $\boldsymbol{\theta}$ . The parametric local likelihood ratio  $\lambda_y(\mathbf{s}, t)$  can be written as

$$\lambda_y(\mathbf{s}, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{s}^T \dot{l}_\theta(\mathbf{Y}_i) + t \dot{l}_\psi(\mathbf{Y}_i)] - [\mathbf{s}^T]^T I [\mathbf{s}^T]/2 + o_p(1)$$

which, under i.i.d. sampling from  $N_p(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})/\psi)$ , converges in distribution to a  $N(-\mathbf{u}^T I \mathbf{u}/2, \mathbf{u}^T I \mathbf{u})$  random variable, where  $\mathbf{u}^T = (\mathbf{s}^T, t)$  and  $I$  is the information matrix for  $(\boldsymbol{\theta}, \psi)$ . The elements of  $I$  needed for the results in this section are

$$\begin{aligned} I_{\psi \boldsymbol{\theta}} &= \{I_{\psi \theta_k}\} = \{-\text{tr}(\mathbf{B} \mathbf{C}_{\theta_k})/(2\psi)\} \\ I_{\psi \psi} &= p/(2\psi^2), \end{aligned}$$

where the fact that  $\mathbf{B}_{\theta_k} \mathbf{C} = -\mathbf{B} \mathbf{C}_{\theta_k}$  has been used in the calculation of  $I_{\psi \boldsymbol{\theta}}$ . If we set  $t = -I_{\psi \psi}^{-1} I_{\psi \boldsymbol{\theta}} \mathbf{s}$ , then  $\lambda_y$  converges in distribution to a  $N(-\mathbf{s}^T I_{\boldsymbol{\theta} \boldsymbol{\theta} \cdot \psi} \mathbf{s}/2, \mathbf{s}^T I_{\boldsymbol{\theta} \boldsymbol{\theta} \cdot \psi} \mathbf{s})$  random variable, where  $I_{\boldsymbol{\theta} \boldsymbol{\theta} \cdot \psi} = I_{\boldsymbol{\theta} \boldsymbol{\theta}} - I_{\boldsymbol{\theta} \psi} I_{\boldsymbol{\theta} \psi}^T / I_{\psi \psi}$  is the information for  $\boldsymbol{\theta}$  in this parametric model.

We will now find conditions on  $\mathbf{C}(\boldsymbol{\theta})$  under which  $\lambda_r(\mathbf{s})$  converges in distribution to the same normal random variable. A candidate rank-measurable approximation to  $\lambda_y(\mathbf{s}, t)$  is given by

$$\lambda_{\hat{y}}(\mathbf{s}, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{s}^T \dot{l}_\theta(\hat{\mathbf{Y}}_i) + t \dot{l}_\psi(\hat{\mathbf{Y}}_i)] - [\mathbf{s}^T]^T I [\mathbf{s}^T]/2.$$

Recall that if for our given  $\mathbf{s}$  and  $\boldsymbol{\theta}$  we can find a  $t$  and  $\psi$  such that  $\lambda_{\hat{y}} - \lambda_y = o_p(1)$ , then the conditions of Theorem 2.4 will be met and the asymptotic distribution of  $\lambda_r(\mathbf{s})$  will be that of  $\lambda_y(\mathbf{s}, t)$ . With this in mind, let  $t = \mathbf{h}^T \mathbf{s}$  for some  $\mathbf{h} \in \mathbb{R}^q$ , and write  $\lambda_y(\mathbf{s}, \mathbf{h}^T \mathbf{s}) \equiv \lambda_y(\mathbf{s})$ . We will find conditions on  $\mathbf{C}(\boldsymbol{\theta})$  such that there exists an  $\mathbf{h}$  for which  $\lambda_{\hat{y}}(\mathbf{s}) - \lambda_y(\mathbf{s}) = o_p(1)$ , and will show that any such  $\mathbf{h}$  must be equal to  $-I_{\psi\psi}^{-1} I_{\psi\boldsymbol{\theta}}$ .

With  $t = \mathbf{h}^T \mathbf{s}$  and  $\psi = 1$ , we have

$$\begin{aligned} \mathbf{s}^T \dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + t \dot{l}_{\psi}(\mathbf{y}) &= \mathbf{s}^T [\dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + \mathbf{h} \dot{l}_{\psi}(\mathbf{y})] \\ &= \sum_{k=1}^q s_k [\dot{l}_{\theta_k}(\mathbf{y}) + h_k \dot{l}_{\psi}(\mathbf{y})] \\ &= - \sum_{k=1}^q s_k \mathbf{y}^T (\mathbf{B}_{\theta_k} + h_k \mathbf{B}) \mathbf{y} / 2 + c(\boldsymbol{\theta}, \mathbf{s}, \mathbf{h}) \\ &= \sum_{k=1}^q s_k \mathbf{y}^T \mathbf{A}_k \mathbf{y} + c(\boldsymbol{\theta}, \mathbf{s}, \mathbf{h}), \end{aligned}$$

where  $\mathbf{A}_k = -(\mathbf{B}_{\theta_k} + h_k \mathbf{B})/2 = (\mathbf{B}\mathbf{C}_{\theta_k}\mathbf{B} - h_k \mathbf{B})/2$  and  $c(\boldsymbol{\theta}, \mathbf{s}, \mathbf{h})$  does not depend on  $\mathbf{y}$ . The difference between  $\lambda_{\hat{y}}$  and  $\lambda_y$  is then

$$\lambda_{\hat{y}} - \lambda_y = \sum_{k=1}^q s_k \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\mathbf{Y}}_i \mathbf{A}_k \hat{\mathbf{Y}}_i - \mathbf{Y}_i \mathbf{A}_k \mathbf{Y}_i \right) + o_p(1).$$

Since  $\mathbf{A}_k$  is symmetric, Theorem 3.2 implies that this difference will converge in probability to zero if the diagonal elements of  $\mathbf{A}_k \mathbf{C}$  are zero for each  $k = 1, \dots, q$ . This condition can equivalently be written as follows:

$$\begin{aligned} \mathbf{0} &= \text{diag}(\mathbf{A}_k \mathbf{C}) \\ &= \text{diag}(\mathbf{B}\mathbf{C}_{\theta_k}\mathbf{B}\mathbf{C} - h_k \mathbf{B}\mathbf{C})/2 \\ &= \text{diag}(\mathbf{B}\mathbf{C}_{\theta_k} - h_k \mathbf{I})/2 \\ h_k \mathbf{1} &= \text{diag}(\mathbf{B}\mathbf{C}_{\theta_k}). \end{aligned}$$

The above condition can only be met if, for each  $k$ , the diagonal elements of  $\mathbf{B}\mathbf{C}_{\theta_k}$  all take on a common value. If they do, then the convergence in probability of  $\lambda_{\hat{y}}(\mathbf{s}, t) - \lambda_y(\mathbf{s}, t)$  to zero can be obtained by setting  $t = \mathbf{h}^T \mathbf{s}$ , where  $h_k = \text{tr}(\mathbf{B}\mathbf{C}_{\theta_k})/p$ .

Note that in this case where  $\psi = 1$ ,  $h_k = \text{tr}(\mathbf{B}\mathbf{C}_{\theta_k})/p = -I_{\psi\psi}^{-1} I_{\psi\theta_k}$ . By setting  $t = -I_{\psi\psi}^{-1} I_{\psi\boldsymbol{\theta}} \mathbf{s} = \mathbf{h}^T \mathbf{s}$ , the value of  $\mathbf{A}_k$  will satisfy the conditions of Theorem 3.2 for each  $k = 1, \dots, q$ . Therefore,  $\lambda_{\hat{y}} - \lambda_y = o_p(1)$  under i.i.d. sampling from  $N_p(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})/\psi)$ , and so the

conditions of Theorem 2.4 are met. The limiting distribution of  $\lambda_r$  under sampling from the copula is therefore the same as that of  $\lambda_y$  under sampling from the corresponding multivariate normal distribution.

**Theorem 4.1.** *Let  $\{\mathbf{C}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q\}$  be a collection of correlation matrices such that  $\mathbf{C}(\boldsymbol{\theta})$  is twice differentiable, and for each  $k$ , the diagonal entries of  $\mathbf{B}\mathbf{C}_{\theta_k}$  are equal to some common value. If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d. from a population with continuous marginal distributions and copula  $\mathbf{C}(\boldsymbol{\theta})$  for some  $\boldsymbol{\theta} \in \Theta$ , then the distribution of the rank likelihood ratio  $\lambda_r(\mathbf{s})$  converges to a  $N(-\mathbf{s}^T I_{\boldsymbol{\theta}\boldsymbol{\theta} \cdot \psi} \mathbf{s} / 2, \mathbf{s}^T I_{\boldsymbol{\theta}\boldsymbol{\theta} \cdot \psi} \mathbf{s})$  distribution, where  $I_{\boldsymbol{\theta}\boldsymbol{\theta} \cdot \psi}$  is the information for  $\boldsymbol{\theta}$  in the normal model with correlation  $\mathbf{C}(\boldsymbol{\theta})$  and equal marginal precisions  $\psi$ .*

### 4.3 Examples

The condition on the diagonal entries of  $\mathbf{B}\mathbf{C}_{\theta_k}$  is satisfied for some well-known models. For example, the one-parameter exchangeable correlation matrix  $\{\mathbf{C}(\theta) : \theta \in [-1, 1]\}$ , for which all off-diagonal elements are equal to  $\theta$ , satisfies the condition. In fact, the condition will be satisfied for any model in which the rows of  $\mathbf{C}(\boldsymbol{\theta})$  are permutations of one another. To see this, note that if  $\mathbf{c}_i$ , the  $i$ th row of  $\mathbf{C}(\boldsymbol{\theta})$ , is a permutation of  $\mathbf{c}_j$ , then  $\mathbf{b}_i$ , the  $i$ th row of  $\mathbf{B}$ , is the same permutation of  $\mathbf{b}_j$ . Therefore  $\mathbf{b}_i^T \mathbf{c}_{\theta_k, i} = \mathbf{b}_j^T \mathbf{c}_{\theta_k, j}$  for each  $i, j$  and  $k$ . Subclasses of such correlation matrices includes circular correlation models, often used for seasonal data [Olkin and Press, 1969, Khattree and Naik, 1994], and any model in which the rows of  $\mathbf{C}$  are permutations of circular matrices.

For illustration, we calculate the information  $I_{\boldsymbol{\theta}\boldsymbol{\theta} \cdot \psi}$  for two one-parameter models satisfying the symmetry condition on  $\mathbf{B}\mathbf{C}_{\theta}$ . For any one-parameter model, the second derivative of the log-likelihood with respect to  $\theta$  is

$$\begin{aligned} \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\mathbf{y}) &= \frac{\partial}{\partial \theta} (\text{tr}(\mathbf{B}_{\theta} \mathbf{C}) - \psi \mathbf{y}^T \mathbf{B}_{\theta} \mathbf{y}) / 2 \\ &= (\text{tr}(\mathbf{B}_{\theta} \mathbf{C}_{\theta}) + \text{tr}(\mathbf{B}_{\theta\theta} \mathbf{C}) - \psi \mathbf{y}^T \mathbf{B}_{\theta\theta} \mathbf{y}) / 2 \\ &= (-\text{tr}(\mathbf{B}\mathbf{C}_{\theta} \mathbf{B}\mathbf{C}_{\theta}) + \text{tr}(\mathbf{B}_{\theta\theta} \mathbf{C}) - \psi \mathbf{y}^T \mathbf{B}_{\theta\theta} \mathbf{y}) / 2 \end{aligned}$$

and so the information for  $\theta$  if  $\psi$  were known is  $I_{\theta\theta} = -E[\ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\mathbf{Y})] = \text{tr}(\mathbf{B}\mathbf{C}_{\theta} \mathbf{B}\mathbf{C}_{\theta}) / 2$ . The information for  $\theta$  when  $\psi$  is unknown is then

$$\begin{aligned} I_{\boldsymbol{\theta}\boldsymbol{\theta} \cdot \psi} &= I_{\theta\theta} - I_{\psi\psi}^2 / I_{\psi\psi} \\ &= \frac{1}{2} [\text{tr}(\mathbf{B}\mathbf{C}_{\theta} \mathbf{B}\mathbf{C}_{\theta}) - \text{tr}(\mathbf{B}\mathbf{C}_{\theta})^2 / p], \end{aligned}$$

so the information loss in going from known to unknown margins is  $I_{\psi\theta}^2 / I_{\psi\psi} = \text{tr}(\mathbf{B}\mathbf{C}_{\theta})^2 / (2p)$ .



Figure 1 plots  $I_{\theta\theta}^{-1}$  and  $I_{\theta\theta\cdot\psi}^{-1}$  for two different copula models in which the symmetry condition on  $\mathbf{BC}_\theta$  is satisfied. The first panel of the figure plots information bounds for the exchangeable correlation model with  $p = 4$ , in which  $\text{Cor}[Y_j, Y_k] = \theta$  for all  $j \neq k$ . For this model, straightforward calculations show that

$$I_{\theta\theta} = (1 - \theta)^{-2} p(p\gamma^2 - 2\gamma + 1)/2, \text{ where } \gamma^{-1} \equiv 1 + (p - 1)\theta,$$

$$I_{\theta\theta\cdot\psi} = \frac{p(p - 1)}{2((p - 1)\theta + 1)^2(1 - \theta)^2},$$

which of course reduces to  $(1 - \theta^2)^{-2}$ , the information bound for the bivariate normal copula model, when  $p = 2$ . Note that the parameter space is  $\Theta = (-(p - 1)^{-1}, 1)$ , and that there is very little information loss for  $\theta < 0$  relative to the loss for  $\theta > 0$ .

The second panel of the figure gives information bounds for a  $4 \times 4$  circular correlation matrix for which the first row is  $\mathbf{c}_1(\theta) = (1, \theta, \theta^2, \theta)$ . For this model, we have

$$I_{\theta\theta} = \frac{4}{(1 - \theta^2)^2} (1 + 2\theta^2),$$

$$I_{\theta\theta\cdot\psi} = \frac{4}{(1 - \theta^2)^2}.$$

Note that the information bounds for this model are symmetric about  $\theta = 0$ .

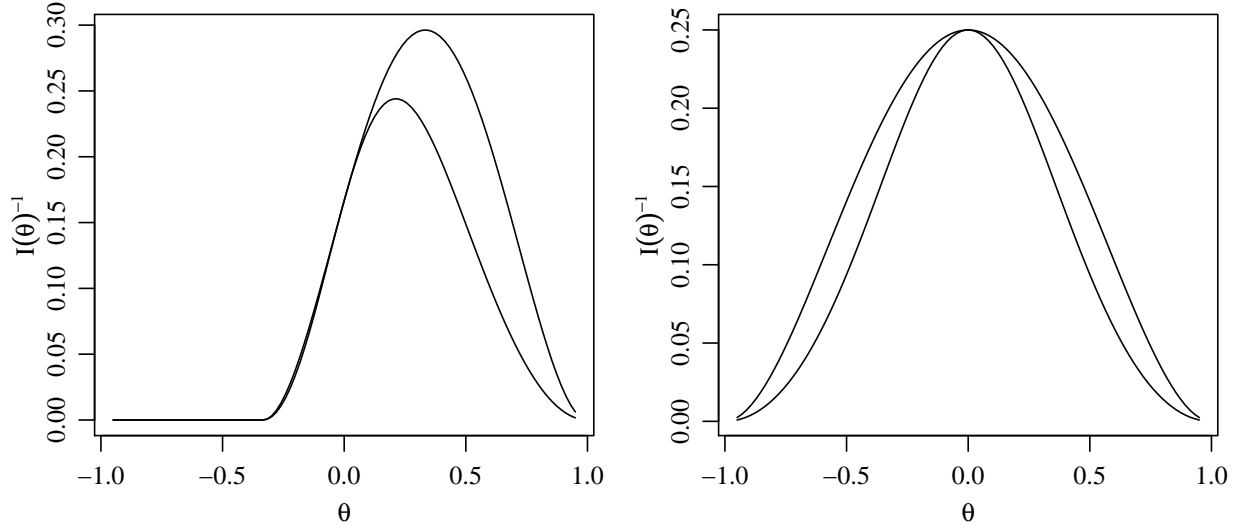


Figure 1: Asymptotic information bounds for copula models with known and equal but unknown margins. The left panel gives  $I_{\theta\theta}^{-1}$  and  $I_{\theta\theta\cdot\psi}^{-1}$  for the  $p = 4$  exchangeable copula model, and the right panel gives bounds for a one parameter circular copula model.

## 5 LAN for general Gaussian copulas

For a copula model with the symmetry in  $\mathbf{BC}_\theta$  described above, the limiting distribution of the rank likelihood ratio is the same as that of the likelihood ratio for the corresponding normal model with common marginal variances, i.e. the model is “symmetric” in the variances. Analogously, one might expect that for a Gaussian copula model lacking this symmetry, the limiting distribution of  $\lambda_r$  might match that of the likelihood ratio for the corresponding normal model with “asymmetric,” i.e. unequal marginal variances. This turns out to be correct, as we now show.

Consider the class of mean-zero multivariate normal models with inverse-covariance matrix  $\text{Var}[\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi}]^{-1} = \mathbf{D}(\boldsymbol{\psi})^{1/2}\mathbf{B}(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\psi})^{1/2}$ , where  $\boldsymbol{\theta} \in \mathbb{R}^q$  and  $\mathbf{D}(\boldsymbol{\psi})$  is the diagonal matrix with diagonal elements  $\boldsymbol{\psi} \in \mathbb{R}^p$ . The log probability density for a member of this class is given by

$$l(\mathbf{y}) = \left( -p \log 2\pi + \sum \log \psi_j + \log |\mathbf{B}| - \mathbf{y}^T \mathbf{D}(\boldsymbol{\psi})^{1/2} \mathbf{B} \mathbf{D}(\boldsymbol{\psi})^{1/2} \mathbf{y} \right) / 2.$$

The log-likelihood derivatives are

$$\begin{aligned} \dot{l}_{\theta_k}(\mathbf{y}) &= [\text{tr}(\mathbf{B}_{\theta_k} \mathbf{C}) - \mathbf{y}^T \mathbf{D}(\boldsymbol{\psi})^{1/2} \mathbf{B}_{\theta_k} \mathbf{D}(\boldsymbol{\psi})^{1/2} \mathbf{y}] / 2 \\ \dot{l}_{\psi_j}(\mathbf{y}) &= [1 - y_j \psi_j^{1/2} \mathbf{b}_j^T \mathbf{D}(\boldsymbol{\psi})^{1/2} \mathbf{y}] / (2\psi_j). \end{aligned}$$

Straightforward calculations show that

$$\begin{aligned} I_{\psi\psi} &= \mathbf{D}(\boldsymbol{\psi})^{-1} (\mathbf{I} + \mathbf{B} \circ \mathbf{C}) \mathbf{D}(\boldsymbol{\psi})^{-1} / 4 \\ I_{\psi\theta_k} &= -\mathbf{D}^{-1}(\boldsymbol{\psi}) \text{diag}(\mathbf{B} \mathbf{C}_{\theta_k}) / 2, \end{aligned}$$

where “ $\circ$ ” is Hadamard product denoting element-wise multiplication. The local log likelihood ratio for this model can be expressed as

$$\lambda_y(\mathbf{s}, \mathbf{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{s}^T \dot{l}_{\boldsymbol{\theta}}(\mathbf{Y}_i) + \mathbf{t}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{Y}_i) - \frac{1}{2} [\mathbf{s}^T \mathbf{t}]^T I [\mathbf{s}^T \mathbf{t}] + o_p(1).$$

As before, we take our rank based approximation  $\lambda_{\hat{y}}$  to be equal to  $\lambda_y$  absent the  $o_p(1)$  term and with each  $\mathbf{Y}_i$  replaced by its approximate normal scores  $\hat{\mathbf{Y}}_i$ . Clearly, we have  $\lambda_{\hat{y}} - \lambda_y = o_p(1)$  if

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{s}^T \dot{l}_{\boldsymbol{\theta}}(\hat{\mathbf{Y}}_i) + \mathbf{t}^T \dot{l}_{\boldsymbol{\psi}}(\hat{\mathbf{Y}}_i)] - [\mathbf{s}^T \dot{l}_{\boldsymbol{\theta}}(\mathbf{Y}_i) + \mathbf{t}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{Y}_i)] = o_p(1).$$

Given  $\boldsymbol{\theta}$  and  $\mathbf{s}$ , we now identify a value of  $\mathbf{t}$  for which the above asymptotic result holds. Let  $\mathbf{t} = \mathbf{H}\mathbf{s}$ , where  $\mathbf{H} \in \mathbb{R}^{p \times q}$ , so that

$$\begin{aligned} \mathbf{s}^T \dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + \mathbf{t}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y}) &= \mathbf{s}^T [\dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + \mathbf{H}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y})] \\ &= \sum_{k=1}^q s_k [\dot{l}_{\theta_k}(\mathbf{y}) + \mathbf{h}_k^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y})], \end{aligned}$$

where  $\{\mathbf{h}_k, k = 1, \dots, q\}$  are the columns of  $\mathbf{H}$ . Now  $\dot{l}_{\theta_k}(\mathbf{y})$  and  $\dot{l}_{\boldsymbol{\psi}}(\mathbf{y})$  are both quadratic in  $\mathbf{y}$ . Evaluating at  $\boldsymbol{\psi} = \mathbf{1}$ , we have  $\dot{l}_{\theta_k}(\mathbf{y}) = [\text{tr}(\mathbf{B}_{\theta_k} \mathbf{C}) - \mathbf{y}^T \mathbf{B}_{\theta_k} \mathbf{y}]/2$  and  $\dot{l}_{\psi_j}(\mathbf{y}) = [1 - y_j \mathbf{b}_j^T \mathbf{y}]/2$ , and so

$$\mathbf{h}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y}) = [\mathbf{h}^T \mathbf{1} - \mathbf{y}^T \mathbf{D}(\mathbf{h}) \mathbf{B} \mathbf{y}]/2,$$

where  $\mathbf{D}(\mathbf{h})$  is the diagonal matrix with elements  $\mathbf{h}$ . Therefore, we can write  $\mathbf{s}^T [\dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + \mathbf{H}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y})]$  as

$$\begin{aligned} \mathbf{s}^T [\dot{l}_{\boldsymbol{\theta}}(\mathbf{y}) + \mathbf{H}^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y})] &= \sum_{k=1}^q s_k [\dot{l}_{\theta_k}(\mathbf{y}) + \mathbf{h}_k^T \dot{l}_{\boldsymbol{\psi}}(\mathbf{y})] \\ &= \left( \sum_{k=1}^q s_k \mathbf{y}^T \mathbf{A}_k \mathbf{y} \right) + c(\mathbf{s}, \mathbf{H}, \boldsymbol{\theta}) \end{aligned}$$

where  $c(\mathbf{s}, \mathbf{H}, \boldsymbol{\theta})$  does not depend on  $\mathbf{y}$ , and  $\mathbf{A}_k$  is given by

$$\mathbf{A}_k = -[\mathbf{B}_{\theta_k} + \mathbf{D}(\mathbf{h}_k) \mathbf{B}]/2.$$

Note that if  $\mathbf{h}_k = h_k \mathbf{1}$ , i.e. all the values are common, then the value of  $\mathbf{A}_k$  reduces to the matrix  $\mathbf{A}_k$  in Section 4.2, in the case of equal marginal variances.

Substituting this representation of  $\mathbf{s}^T \dot{l}_{\boldsymbol{\theta}} + \mathbf{t}^T \dot{l}_{\boldsymbol{\psi}}$  into  $\lambda_{\hat{y}}$  and  $\lambda_y$  gives

$$\lambda_{\hat{y}} - \lambda_y = \sum_{k=1}^q s_k \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\mathbf{Y}}_i \mathbf{A}_k \hat{\mathbf{Y}}_i - \mathbf{Y}_i \mathbf{A}_k \mathbf{Y}_i \right) + o_p(1).$$

Theorem 3.2 implies that this difference will converge in probability to zero if the diagonal elements of  $(\mathbf{A}_k + \mathbf{A}_k^T) \mathbf{C}$  are zero for each  $k = 1, \dots, q$ . The value of  $(\mathbf{A}_k + \mathbf{A}_k^T) \mathbf{C}$  can be calculated as

$$\begin{aligned} 2(\mathbf{A}_k + \mathbf{A}_k^T) \mathbf{C} &= -2 \times \mathbf{B}_{\theta_k} \mathbf{C} - \mathbf{D}(\mathbf{h}_k) \mathbf{B} \mathbf{C} - \mathbf{B} \mathbf{D}(\mathbf{h}_k) \mathbf{C} \\ &= 2 \times \mathbf{B} \mathbf{C}_{\theta_k} - (\mathbf{D}(\mathbf{h}_k) + \mathbf{B} \mathbf{D}(\mathbf{h}_k) \mathbf{C}). \end{aligned}$$

The vector  $\text{diag}(\mathbf{D}(\mathbf{h}_k) + \mathbf{B}\mathbf{D}(\mathbf{h}_k)\mathbf{C})$  can be written as

$$\text{diag}(\mathbf{D}(\mathbf{h}_k) + \mathbf{B}\mathbf{D}(\mathbf{h}_k)\mathbf{C}) = \begin{pmatrix} h_{k1} + \mathbf{h}_k^T(\mathbf{b}_1 \circ \mathbf{c}_1) \\ \vdots \\ h_{kp} + \mathbf{h}_k^T(\mathbf{b}_p \circ \mathbf{c}_p) \end{pmatrix} = (\mathbf{I} + \mathbf{B} \circ \mathbf{C})\mathbf{h}_k,$$

and so our condition on  $\mathbf{h}_k$  becomes

$$\begin{aligned} (\mathbf{I} + \mathbf{B} \circ \mathbf{C})\mathbf{h}_k &= 2 \times \text{diag}(\mathbf{B}\mathbf{C}_{\theta_k}) \\ \mathbf{h}_k &= 2(\mathbf{I} + \mathbf{B} \circ \mathbf{C})^{-1} \text{diag}(\mathbf{B}\mathbf{C}_{\theta_k}) \\ &= -I_{\psi\psi}^{-1} I_{\theta_k\psi}. \end{aligned}$$

This result allows us to find the asymptotic distribution of the local rank likelihood ratio  $\lambda_r(\mathbf{s})$  for any smoothly parameterized normal copula model with continuous margins. Given such a copula model, we can form the local log likelihood ratio of the corresponding multivariate normal model with unequal variances,  $\lambda_y(\mathbf{s}, \mathbf{t})$ . If we set  $\mathbf{t} = -I_{\psi\psi}^{-1} I_{\psi\theta} \mathbf{s}$  then by Theorem 3.2 we have  $\lambda_{\hat{y}}(\mathbf{s}, \mathbf{t}) - \lambda_y(\mathbf{s}, \mathbf{t}) = o_p(1)$  under i.i.d. sampling from  $N(\mathbf{0}, \mathbf{C}(\theta))$ . By Theorem 2.4, the limiting distribution of  $\lambda_r(\mathbf{s})$  under i.i.d. sampling from the corresponding copula is therefore  $N(-\mathbf{s}^T I_{\theta\theta\cdot\psi} \mathbf{s} / 2, \mathbf{s}^T I_{\theta\theta\cdot\psi} \mathbf{s})$ , where  $I_{\theta\theta\cdot\psi} = I_{\theta\theta} - I_{\theta\psi}^T I_{\psi\psi}^{-1} I_{\theta\psi}$  is the information for  $\theta$  in the parametric normal model.

**Theorem 5.1.** *Let  $\{\mathbf{C}(\theta) : \theta \in \Theta \subset \mathbb{R}^q\}$  be a collection of correlation matrices such that  $\mathbf{C}(\theta)$  is twice differentiable. If  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d. from a population with continuous marginal distributions and copula  $\mathbf{C}(\theta)$  for some  $\theta \in \Theta$ , then the distribution of the rank likelihood ratio  $\lambda_r(\mathbf{s})$  converges to a  $N(-\mathbf{s}^T I_{\theta\theta\cdot\psi} \mathbf{s} / 2, \mathbf{s}^T I_{\theta\theta\cdot\psi} \mathbf{s})$  distribution, where  $I_{\theta\theta\cdot\psi}$  is the information for  $\theta$  in the normal model with correlation  $\mathbf{C}(\theta)$  and marginal precisions  $\psi$ .*

This theorem generalizes the result of Theorem 4.1 regarding models for which the diagonal elements of  $\mathbf{B}\mathbf{C}_{\theta_k}$  are equal for each  $k$ . Under that condition, the information matrix  $I_{\theta\theta\cdot\psi}$  for  $\theta$  when the marginal variances are equal is the same as  $I_{\theta\theta\cdot\psi}$ , the information for  $\theta$  when the variances are unequal.

## 5.1 Example: First order autoregressive copula

The first-order autoregressive correlation model can be parameterized as  $\mathbf{C}(\theta) = \{c_{j,k}(\theta) = \theta^{|j-k|}\}$ . This simple one-parameter model does not satisfy the conditions of Theorem 4.1, and so the limiting distribution of the rank likelihood in this case is equal to that of the likelihood

for the normal model with unequal marginal variances. For illustration, we compute for this model  $I_{\theta\theta}$ ,  $I_{\theta\theta\cdot\psi}$  and  $I_{\theta\theta\cdot\psi}$ , the information functions for  $\theta$  under known, unknown but equal, and unknown and unequal marginal variances.

For a generic one-parameter Gaussian copula, let  $I_{\theta\psi}$  be the off-diagonal block of the information matrix in the unequal variance model, and let  $I_{\theta\psi}$  be the corresponding element of the information matrix in the equal variance model. Under  $\psi = \mathbf{1}$  and  $\psi = 1$  respectively, we have  $I_{\theta\psi} = -\text{diag}(\mathbf{BC}_\theta)/2$  and  $I_{\theta\psi} = -\text{tr}(\mathbf{BC}_\theta)/2 = I_{\theta\psi}^T \mathbf{1}$ . Letting  $\mathbf{d} = \text{diag}(\mathbf{BC}_\theta)$ , the information in  $\theta$  under these two models can then be written as

$$\begin{aligned} I_{\theta\cdot\psi} &= I_{\theta\theta} - \mathbf{d}^T \begin{bmatrix} \mathbf{1}\mathbf{1}^T/(2p) \\ \end{bmatrix} \mathbf{d} \\ I_{\theta\cdot\psi} &= I_{\theta\theta} - \mathbf{d}^T \begin{bmatrix} (\mathbf{B} \circ \mathbf{C} + \mathbf{I})^{-1} \\ \end{bmatrix} \mathbf{d}. \end{aligned}$$

We have  $I_{\theta\cdot\psi} \geq I_{\theta\cdot\psi}$  with equality if the conditions of Theorem 4.1 are met, i.e. the diagonal elements of  $\mathbf{BC}_\theta$  are all equal.

The first panel of Figure 2 plots  $I_{\theta\theta}$ ,  $I_{\theta\cdot\psi}$  and  $I_{\theta\cdot\psi}$  for the first-order autoregressive model with  $p = 4$ . Note that the information loss in going from equal to unequal variances is quite small, as compared to going from known to unknown marginal variances. The second panel of the figure highlights these differences.

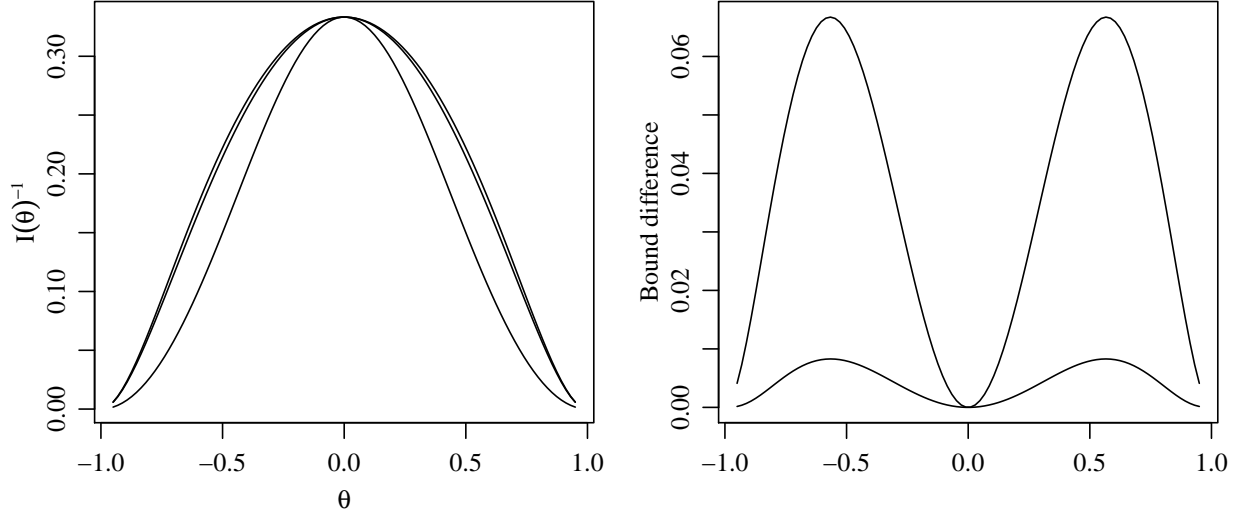


Figure 2: Asymptotic information bounds for the first-order autoregressive normal copula model with known, equal but unknown margins, and unequal and unknown margins. The right panel gives the bounds for the three cases, and the second panel gives the differences  $I_{\theta\theta\cdot\psi}^{-1} - I_{\theta\theta}^{-1}$  and  $I_{\theta\theta\cdot\psi}^{-1} - I_{\theta\theta\cdot\psi}^{-1}$ , the latter one being the smaller of the two.

## 6 Discussion

The partial sufficiency of the multivariate ranks in semiparametric copulas models suggests the existence of asymptotically efficient rank-based estimators of copula parameters. For the one-parameter bivariate Gaussian copula model, the rank-based pseudo-likelihood estimator of Genest et al. [1995] is asymptotically equivalent to the normal scores correlation coefficient, which Klaassen and Wellner [1997] showed to be asymptotically efficient. For other copula models, the existence of efficient rank-based estimators is an open question. Genest and Werker [2002] showed with a non-Gaussian example that the pseudo-likelihood estimator is not generally asymptotically efficient. However, this does not rule out the possibility that other rank-based estimators are asymptotically efficient, or that pseudo-likelihood estimators would be efficient for general Gaussian copula models.

A natural candidate for an efficient rank-based estimator is the maximizer of the rank likelihood. However, whereas the pseudo-likelihood is a very explicit function of the copula density (making optimization and asymptotic analysis tractable), the rank likelihood involves a multivariate integral over a set of order constraints, the number of which grows with the sample size. While rank likelihood copula estimation can be made feasible with standard Markov chain Monte Carlo integration techniques [Hoff, 2007], an asymptotic analysis of the rank likelihood seems to require techniques tailored to this particular problem.

In this article, we have shown that the existence of a sufficiently accurate rank measurable approximation to a parametric submodel implies the local asymptotic normality of the rank likelihood. We have also shown that such approximations exist for every smoothly parameterized Gaussian copula model. For such a copula model, the asymptotic information bound implied by the rank likelihood matches that of the corresponding parametric multivariate normal submodel. This result suggests the possibility of asymptotically efficient rank-based estimators for Gaussian copula models: Generally speaking, the information  $I_r$  based on the ranks is less than or equal to the semiparametric information  $I_f$  based on the full data, as the ranks are functions of the full data [Le Cam and Yang, 1988]. Furthermore, the semiparametric information based on the full data is less than or equal to  $I_p$ , the infimum of information functions over all parametric submodels, and so  $I_r \leq I_f \leq I_p$  in general. On the other hand, for Gaussian copula models we have shown that  $I_r$  is equal to the information for a particular parametric submodel, the corresponding multivariate normal model. This implies that for a given Gaussian copula model, the corresponding multivariate normal model is least favorable, that  $I_r = I_p$  and therefore  $I_r = I_f = I_p$ . Based on this result,

we conjecture that maximum likelihood estimators based on rank likelihoods are asymptotically efficient for Gaussian copula models, and possibly more generally whenever information bounds based on the complete data for the semiparametric model in question exist.

## References

- G. A. Barnard. Logical aspects of the fiducial argument. *Bull. Inst. Internat. Statist.*, 40: 870–883, 1963.
- Janet M. Begun, W. J. Hall, Wei-Min Huang, and Jon A. Wellner. Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.*, 11(2):432–452, 1983. ISSN 0090-5364. doi: 10.1214/aos/1176346151. URL <http://dx.doi.org/10.1214/aos/1176346151>.
- P. J. Bickel and Y. Ritov. Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for Lucien Le Cam*, pages 43–54. Springer, New York, 1997.
- Peter J. Bickel, Chris A. J. Klaassen, Ya’acov Ritov, and Jon A. Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1993. ISBN 0-8018-4541-6.
- Xiaohong Chen, Yanqin Fan, and Viktor Tsyrennikov. Efficient estimation of semiparametric multivariate copula models. *J. Amer. Statist. Assoc.*, 101(475):1228–1240, 2006. ISSN 0162-1459. doi: 10.1198/016214506000000311. URL <http://dx.doi.org/10.1198/016214506000000311>.
- Sungsub Choi, W. J. Hall, and Anton Schick. Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, 24(2):841–861, 1996. ISSN 0090-5364. doi: 10.1214/aos/1032894469. URL <http://dx.doi.org/10.1214/aos/1032894469>.
- T. de Wet and J. H. Venter. Asymptotic distributions of certain test criteria of normality. *South African Statist. J.*, 6:135–149, 1972. ISSN 0038-271X.
- C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995. ISSN 0006-3444.

- Christian Genest and Bas J. M. Werker. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In *Distributions with given marginals and statistical modelling*, pages 103–112. Kluwer Acad. Publ., Dordrecht, 2002.
- W. J. Hall and R. M. Loynes. On the concept of contiguity. *Ann. Probability*, 5(2):278–282, 1977.
- Peter D. Hoff. Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.*, 1(1):265–283, 2007. ISSN 1932-6157.
- Peter D. Hoff. Rank likelihood estimation for continuous and discrete data. *ISBA Bulletin*, 15(1):8–10, 2008. URL <http://bayesian.org/sites/default/files/fm/bulletins/0803.pdf>.
- Ravindra Khattree and Dayanand N. Naik. Estimation of interclass correlation under circular covariance. *Biometrika*, 81(3):612–617, 1994. ISSN 0006-3444. doi: 10.1093/biomet/81.3.612. URL <http://dx.doi.org/10.1093/biomet/81.3.612>.
- Chris A. J. Klaassen and Jon A. Wellner. Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 3(1):55–77, 1997. ISSN 1350-7265.
- Lucien Le Cam and Grace L. Yang. On the preservation of local asymptotic normality under information loss. *Ann. Statist.*, 16(2):483–520, 1988. ISSN 0090-5364. doi: 10.1214/aos/1176350817. URL <http://dx.doi.org/10.1214/aos/1176350817>.
- I. Olkin and S. J. Press. Testing and estimation for a circular stationary model. *Ann. Math. Statist.*, 40:1358–1373, 1969. ISSN 0003-4851.
- A. N. Pettitt. Inference for the linear model using a likelihood based on ranks. *J. Roy. Statist. Soc. Ser. B*, 44(2):234–243, 1982. ISSN 0035-9246.
- M. Rémon. On a concept of partial sufficiency:  $L$ -sufficiency. *Internat. Statist. Rev.*, 52(2):127–135, 1984. ISSN 0306-7734.
- Thomas A. Severini. *Likelihood methods in statistics*, volume 22 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2000. ISBN 0-19-850650-3.
- M. Sklar. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.